TONG ZHOU

zhou.tong1@northeastern.edu
Google Scholar | LinkedIn | Homepage

RESEARCH INTERESTS

My research focuses on **securing AI systems against adversarial threats** throughout the model lifecycle, with an emphasis on **intellectual property protection**, **trustworthy deployment**, and **responsible generative AI**. I develop practical and effective defenses to make AI systems verifiable, robust, and resistant to misuse.

EDUCATION

Northeastern University, Boston, MA, USA — Sep. 2021 – present
Ph.D. in Electrical & Computer Engineering
Advisor: Prof. Xiaolin Xu

University of Michigan, Ann Arbor, MI, USA — Sep. 2019 – Apr. 2021
M.S. in Electrical & Computer Engineering

Xidian University, Xi'an, Shaanxi, China — Sep. 2015 – Jul. 2019
B.S. in Electrical Engineering

SELECTED AWARDS

NeurIPS Scholar Award 2024

ICML Travel Grant 2023

COE Outstanding Graduate Student Award, Northeastern University 2023

IEEE/ACM William J. McCalla ICCAD Best Paper Nomination 2022

COE Dean's Fellowship Award, Northeastern University 2021

Outstanding Graduate Award , Xidian University 2019

First Prize Scholarship , Xidian University 2016 - 2018

PUBLICATIONS (*indicates equal contribution)

Conference Proceedings

[C10] Probe-Me-Not: Protecting Pre-trained Encoders from Malicious Probing
Duyi Ding, **Tong Zhou**, Lili Su, Adam Ding, Xiaolin Xu, and Yunsi Fei
In Proceedings of the 2025 Annual Network and Distributed System Security Symposium (**NDSS**), 2025.

[C9] Bileve: Securing Text Provenance in Large Language Models Against Spoofing with Bi-level Signature
Tong Zhou, Xuandong Zhao, Xiaolin Xu, and Shaolei Ren
The Thirty-eighth Annual Conference on Neural Information Processing Systems (**NeurIPS**), 2024.

[C8] AdaPI: Facilitating Dnn Model Adaptivity For Efficient Private Inference in Edge Computing
Tong Zhou*, Jiahui Zhao*, Yukui Luo, Xi Xie, Wujie Wen, Caiwen Ding, Xiaolin Xu
IEEE/ACM International Conference on Computer-Aided Design (**ICCAD**), 2024

[C7] TBNet: A Neural Architectural Defense Framework Facilitating DNN Model Protection in Trusted Execution Environments
Ziyu Liu, **Tong Zhou**, Yukui Luo, Xiaolin Xu
In Proceedings of the 61st ACM/IEEE Design Automation Conference (**DAC**), 2024

[C6] ArchLock: Locking DNN Transferability at the Architecture Level with a Zero-Cost Binary Predictor
Tong Zhou, Shaolei Ren, and Xiaolin Xu
The Twelfth International Conference on Learning Representations (**ICLR**), 2024.

[C5] MirrorNet: A TEE-Friendly Framework for Secure On-device DNN Inference
Ziyu Liu, Yukui Luo, Shijin Duan, **Tong Zhou** and Xiaolin Xu
IEEE/ACM International Conference on Computer-Aided Design (**ICCAD**), 2023.

[C4] AutoReP: Automatic ReLU Replacement for Fast Private Network Inference
Tong Zhou*, Hongwu Peng*, Shaoyi Huang*, Xiaolin Xu, Caiwen Ding, *et al*.
International Conference on Computer Vision (**ICCV**), 2023.

[C3] NNSplitter: An Active Defense Solution to DNN Model via Automated Weight Obfuscation
Tong Zhou, Yukui Luo, Shaolei Ren, Xiaolin Xu
International Conference on Machine Learning (**ICML**), 2023.

[C2] ObfuNAS: A Neural Architecture Search-based DNN Obfuscation Approach
Tong Zhou, Shaolei Ren, Xiaolin Xu
IEEE/ACM International Conference On Computer Aided Design (**ICCAD**), 2022.
Best Paper Nomination

[C1] Deep neural network security from a hardware perspective
Tong Zhou, Yuheng Zhang, Shijin Duan, Yukui Luo, Xiaolin Xu
IEEE/ACM International Symposium on Nanoscale Architectures (**NANOARCH**), 2021

Journal Papers

[J1] Neural architecture search for adversarial robustness via learnable pruning
Yize Li, Pu Zhao, Ruyi Ding, **Tong Zhou**, Yunsi Fei, Xiaolin Xu, Xue Lin
Frontiers in High Performance Computing, 2024

Workshop Papers

[W1] ProDiF: Protecting Domain-Invariant Features to Secure Pre-Trained Models Against Extraction
Tong Zhou, Shijin Duan, Gaowen Liu, Charles Fleming, Shaolei Ren, Xiaolin Xu, *et al*.
ICLR Workshop on Neural Network Weights as a New Data Modality, 2025

Professional Experience

Applied Scientist Intern @ Microsoft Redmond, WA
Manager: Dr. Tao Ge *Jun. 2025 – present*

This project focuses on personalized long-form text generation, aiming to produce coherent and stylistically consistent content tailored to individual users.

Applied Scientist Intern @ Amazon San Deigo, CA
Manager: Dr. Tao Yuan *May 2024 – Aug. 2024*
Developed a unified model to improve account takeover detection by leveraging multiple data sources. (Accepted to **Amazon Machine Learning Conference Workshop 2024**).

Research Assistant @ Jiande Chen's Lab Ann Arbor, MI
Advisor: Prof. Jiande Chen *Nov. 2020 – Apr. 2021*
Developed deep learning models for feature extraction from electrocardiogram data to detect food intake phases, aiming to assist in treating obesity and diabetes.

Research Assistant @ Laboratory of Integrated Brain Imaging Ann Arbor, MI
Advisor: Prof. Zhongming Liu *May 2020 – Oct. 2020*
Enhanced segmentation performance for Transmission Electron Microscopy (TEM) images by integrating a self-attention mechanism into the U-Net architecture.

Teaching Experience

Teaching Assistant Northeastern University
EECE 2311: Lab for Digital Design Fall 2024

- Led lab sections on digital logic design and FPGA development.
- Guided students through circuit simulation, synthesis, and debugging on hardware.
- Graded lab reports and assisted with conceptual reinforcement during office hours.

Teaching Assistant Northeastern University
EECE 7390: Computer Hardware Security Spring 2025

- Provided pedagogical support for a graduate-level course on hardware security.
- Collaborated with the instructor to address common learning challenges.
- Graded assignments and exams with consistency and timely feedback.

Academic Services

Conference Reviewer

The International Conference on Machine Learning (**ICML**), 2025
The International Conference on Learning Representations (**ICLR**), 2025
The International Conference on Neural Information Processing Systems (**NeurIPS**), 2024, 2025
The International Conference on Artificial Intelligence and Statistics (**AISTATS**), 2025
IEEE International Symposium on Hardware Oriented Security and Trust (**HOST**), 2023
IEEE International Conference on Computer Design (**ICCD**), 2022

Journal Reviewer

IEEE Transactions on Information Forensics and Security (**TIFS**)
IEEE Systems Journal (**ISJ**)

Volunteer

The International Conference on Machine Learning (**ICML**), 2023
New England Hardware Security Workshop (**NEHWS**), 2023

Invited Talks

- **Anti-forgery watermarks for AI-generated contents**
  UMass Dartmouth CIS Seminar Apr. 2025